
CAPSTONE PROJECT

FACULTY DIGITAL PROFILE BUILDER

Presented By:

**1. Joydip Karmakar - Government College of Engineering and
Textile Technology Serampore (GCETTS) - CSE**

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Academic institutions require **accurate, updated digital profiles of faculty members** for:

- Accreditation bodies
- Conference submissions
- Institutional websites
- Funding agencies

Currently:

- Profile creation is manual and repetitive.
- Data is scattered across CVs, certificates, papers, and reports.
- Extracting structured information from unstructured documents is difficult and time-consuming.

PROPOSED SOLUTION

- We propose building an **AI Agent** hosted on **IBM Cloud** that:
- Uses **watsonx.ai** with **Retrieval-Augmented Generation (RAG)** to extract structured information from unstructured documents.
- Processes inputs like:
 - CV PDFs / DOCX
 - Certificate images / documents
 - Publication lists / grant proposals
- Outputs:
 - Structured JSON containing profile data.
 - Exportable formats for further use in institutional systems.
- Agent flow:
 - → Input Document(s) → RAG-based Extraction → JSON Profile → Stored on IBM Cloud Storage / Database

SYSTEM APPROACH

Component	Technology / Tool Used
Cloud Platform	IBM Cloud
AI Model	watsonx.ai Foundation Model (Fine-tuned)
Retrieval Component	IBM Cloud Object Storage + Vector DB (optional)
Workflow Orchestration	IBM Cloud Functions / IBM Watson Studio Pipelines
Data Storage	IBM Cloud Object Storage / Cloudant DB
Document Processing	PyPDF2, docx2txt, Tesseract OCR (for images)
Output Format	JSON, CSV

ALGORITHM & DEPLOYMENT

Algorithm

- Input Collection
 - Upload CV/documents to IBM Cloud Object Storage.
- Pre-processing
 - Convert PDFs, DOCX, images to raw text.
- RAG-based AI Processing
 - Retrieve context → Generate structured data (Name, Department, Education, Certifications, Patents, Grants, Publications, Work Experience) via watsonx.ai.
- Post-processing
 - Clean, format, and structure data into JSON.
- Storage & Output
 - Save JSON to IBM Cloud Storage / Cloudant; expose via API.

Deployment

- AI Agent on IBM Cloud Functions

RESULT

Build

AI

Model: granite-3-3-8b-instruct

<

Setup

Configuration

Framework

LangGraph

Architecture

ReAct

Instructions

Advanced configuration

You are a helpful assistant designed to extract comprehensive academic profile information for a faculty member from provided text.

Your task is to identify and extract the following details and present them in a structured JSON format. If a piece of information is not explicitly found or is empty, provide an empty

Knowledge

Tools

Add a tool

Create custom tool

Added tools (3)

Agent preview

JK

You: 11:59 AM

Curriculum Vitae Name: Dr. Joydip Karmakar Designation: Assistant Professor Department: Computer Science and Engineering Email: joydip.karmakar@university.edu Qualifications Ph.D. in Artificial Intelligence, IIT Kharagpur, 2020 M.Tech in Computer Science, NIT Trichy, 2015 B.Tech in CSE, Jadavpur University, 2013 Publications "AI in Healthcare Systems" – IEEE, 2021 "Deep Learning for Medical Imaging" – Springer, 2022 FDPs Attended 5-Day AICTE Sponsored FDP on Machine Learning – 2023 7-Day FDP on Data Science Tools – 2022 Awards Best Teacher Award – 2023 Young Researcher Award – 2021

BuilderAgent 11:59 AM

{ "name": "Dr. Joydip Karmakar", "designation": "Assistant Professor", "department": "Computer Science and Engineering", "qualifications": [{ "degree": "Ph.D. in Artificial Intelligence", "university": "IIT Kharagpur", "year": 2020 }, { "degree": "M.Tech in Computer Science", "university": "NIT Trichy", "year": 2015 }, { "degree": "B.Tech in CSE", "university": "Jadavpur University", "year": 2013 }], "publications": [{ "title": "AI in Healthcare Systems", "journal": "IEEE", "year": 2021 }, { "title": "Deep Learning for Medical Imaging", "journal": "Springer", "year": 2022 }], "fdps": [{ "name": "5-Day AICTE Sponsored FDP on Machine Learning", "organizer": "NA", "year": 2023 }, { "name": "7-Day FDP on Data Science Tools", "organizer": "NA", "year": 2022 }], "awards": [{ "name": "Best Teacher Award", "awarding_body": "NA", "year": 2023 }, { "name": "Young Researcher Award", "awarding_body": "NA", "year": 2021 }], "research_interests": "Artificial Intelligence, Machine Learning, Medical Imaging" }

► How did I get this answer?

Type something...

CONCLUSION

- The Faculty Digital Profile Builder AI Agent automates the extraction of academic profile data.
- Saves manual effort for faculty and administrative staff.
- Ensures consistency and completeness of digital profiles.
- Hosted entirely on IBM Cloud leveraging watsonx.ai.

FUTURE SCOPE

- Support multi-language document extraction.
- Enable real-time batch processing for entire department profiles.
- Integrate with university ERP systems.
- Add automatic periodic monitoring of new publications (via APIs like Scopus, PubMed).
- Develop RAG pipelines fine-tuned on academic document structures.
- Develop a web solution.

REFERENCES

- IBM watsonx.ai Documentation
- IBM Cloud Object Storage Documentation
- IBM Cloudant Documentation
- IBM Cloud Functions Documentation
- PyPDF2, docx2txt, Tesseract OCR
- "Retrieval-Augmented Generation" Research Papers
- LangChain / Huggingface Transformers

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



JOYDIP KARMAKAR

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence



Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/876e6c7a-1c96-411a-8c2e-0e26ef11b671>



IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



JOYDIP KARMAKAR

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



Issued on: Jul 18, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/ed72e7c6-15b9-45e7-82be-c12ff3d1465c>



IBM CERTIFICATIONS

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to
JOYDIP KARMAKAR

for the completion of

**Lab: Retrieval Augmented Generation with
LangChain**

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT)

Learning hours: 20 mins



THANK YOU